AMENDMENTS TO THE SPECIFICATION

Please replace the paragraph beginning at page 10, line 2 with the following amended paragraph:

For example, the word "data" (with a leading and trailing blank space) can be parsed into the following set of bi-grams: _d, da, at, ta, and a_; and tri-grams: _da, dat, ata, and ta_; and quad-grams: _dat, data, ata_. Generally, a word of length "k", padded with a preceding and trailing blank, will have k-n+3 consecutive overlapping n-grams--k+1 bi-grams, k tri-grams, k-1 quad-grams, and so on. Other types of n-grams that can alternatively or conjointly be used by this method such as anchored n-grams or replacement-type n-grams are described below. Upon parsing the textual passage into a plurality of n-grams 204, the total number of resulting n-grams is calculated and stored 206. One such method of calculating and storing the number of n-grams is disclosed in U.S. Pat. No. 5,062,143, which is hereby incorporated by reference.

Please replace the paragraph beginning at page 11, line 7 with the following amended paragraph:

The frequency with which each n-gram appears in the n-gram language database is thereafter divided by the total number of n-grams in the n-gram language database 214. The resulting quotient is equal to the n-gram's initial ~~weighing factor~~ weighting factor 222. Thus, an initial weighting factor is assigned to each parsed n-gram, as that n-gram relates to a particular language. In order to assign another initial weighting factor to that same n-gram, as the n-gram relates to other languages, the parsed n-gram is compared to another language database that includes n-grams representative of that other language. That is, the process of steps 208, 210, 212, 214 and 216 is repeated for each language with which the n-gram is compared. Parsed n-grams can be compared to all relevant and/or available language databases such that each n-gram is individually compared to all language databases sequentially or the parsed n-grams can be sequentially compared to the language databases as an entire group.

Please replace the paragraph beginning at page 11, line 31 with the following amended paragraph:

As discussed above with respect to step 216, the method of the present invention determines whether each n-gram is present in each particular language database. The number of language databases, within which each n-gram is present, is tabulated and stored 220. The weighting factor for each n-gram that is present in more than one language database is adjusted by multiplying the initial weighting factor and the inverse of the number of databases within which the corresponding n-gram is found 218. In other words, the adjusted weighting factor is equal to the initial weighting factor divided by the number of language databases containing the corresponding n-gram. The adjusted weighting 224 for each n-gram, per language, is summed together providing a passage weight for each language. If the same n-gram appears more than once in a text passage, each instance contributes the adjusted weighting for the n-gram to the sum. The language that has the highest passage weight for the text passage is chosen as the language for the passage. Since each language has a passage weight calculated by this method, it is also possible to rank the possible languages that a text passage may be in. For example, it could be that the text passage has a text weight of 2.29504 for French, of 0.99289 for Spanish, and of 0.843778 for Portuguese, etc. By further comparison of these passage weights, it might be possible to give a level of confidence in the language identification obtained. For example, if the difference between passage weights between the two highest ranked languages was very small, the system might indicate that the text may be one of two languages.